# The state of weather data infrastructure

## White paper

ODI-WP-2017-003
April 2017



Open Data Institute

# Table of contents




**Authors:** Leigh Dodds, Alex Longden, Simon McLellan, Amanda Smith
**Editing:** Anna Scott
**Design:** Christie Brewster

# Executive summary

'Data infrastructure' consists of a combination of data assets, technology, processes and organisations.[1] In this report we use this definition to explore how the ways in which weather and climate data are accessed, used and shared are evolving.

Reviewing weather data's importance, and the steps in its value chain, we draw attention to:

- the evolving role of global, regional and national meteorological services in collecting observations and generating forecasts
- the technology trends creating new data assets along with new big data challenges
- the growing need for investment in technical infrastructure and supercomputing resources to drive improvements in forecasting, and the pressure this puts on ensuring sustainable access to data

Analysing key technology trends and their impacts is important to understanding how weather data infrastructure is evolving. In studying these drivers we hope to encourage discussion around how our global weather data infrastructure can be sustainable and continue to bring significant value to society, while remaining as open as possible.

1 See: 'What is data infrastructure?' at: https://theodi.org/what-is-data-infrastructure.

# Introduction

From its early beginnings over 150 years ago, weather forecasting at the Met Office has been driven by data. Simple observations recorded and used to hand-plot synoptic charts have been exchanged for the billions of observations received and handled every day, mainly from satellites but also from weather stations, radar, ocean buoys, planes, shipping and the public.

However, the huge amounts of data collected, modelled and processed by the supercomputer continues to increase exponentially. This creates a challenge to send weather data to a growing number of consumers via a variety of applications and services, on time and in full.

In this paper, the Open Data Institute (ODI) and the Met Office explore the current state of weather data infrastructure. By looking at the different aspects of the data infrastructure used to collect, access and share weather and climate information, we can better understand challenges.

We hope that by highlighting these challenges we can engage a wide community in agreeing the best way forward.

# The importance of weather data

Like geographical data, information about the weather and climate is of critical importance to many popular applications. Its use benefits society in general as well as many different commercial activities. Weather and climate data can help:

- reduce risks relating to climate- and weather-related disasters
- ensure safe transport of people and freight
- create resilience in the face of climate change and variation
- support sustainable use of natural resources
- drive economic growth across a number of industries

Reports by the ODI have highlighted the importance of weather data in addressing a number of global development changes (Atz et al, 2015) and its role in supporting our global food system (Carolan et al, 2015). The Met Office has estimated that in the UK alone the Public Weather Service contributes more than £1 billion to the national economy.[2] A review by the US

2 See: 'How valuable is the Met Office?' at: http://www.metoffice.gov.uk/about-us/what/pws/value.

Economics and Statistics Administration suggests that US weather forecasts can be valued at $31.5 billion dollars per year (Callen, 2014).

These valuations reflect the fact that all of us consume weather information as part of our daily lives, while organisations large and small are using the same information to make decisions about how they operate and grow. A robust weather data infrastructure is essential to preserving people's safety and for ensuring healthy global, national and regional economies.

There are opportunities to unlock even more value from weather data. With increasing accuracy it becomes possible to create new products and services that are tailored for specific sectors as well as the general public. For example, fine-grained localised forecasting could help route air traffic or guide construction projects. Achieving these benefits requires our weather data infrastructure to adapt and grow.

## The weather data value chain

There are a number of steps involved in drawing value from weather data. At each stage, a number of organisations are involved. National Meteorological Services (NMSs) play a pivotal role in collecting, generating and sharing information as part of their mandate. But achieving this means collaborating with a variety of organisations.

The key stages of the weather data value chain are as follows:

- **Monitoring and observation** of the weather and environment, e.g. by NMSs.
- **Numerical weather prediction (NWP) and climate modelling** carried out by NMSs to create global, regional and limited area weather forecasts. Private companies are growing their presence in the market and challenging the traditional role of NMSs to provide forecasts to the public, by statistically blending data from NMS models to create their own forecast models, for example. Other companies providing data via online channels and/or apps include The Weather Company, Accuweather or the Climate Corporation.
- **Communication and dissemination** of forecasts by news, NMS and media organisations like the BBC, Yahoo and Google, or within consumer-targeted mobile and web applications.
- **Decision making** by individuals and businesses across a variety of sectors, which draws on weather data and reporting.

Trust, accuracy and reliability are important parts of this overall value chain.

## What are the key elements of our global weather data infrastructure?

The Open Data Institute defines data infrastructure as consisting of:

- **data assets**, e.g. datasets collected and created, and analysis carried out on them
- the **organisations** that operate and maintain those assets and technologies
- the **technologies and processes** that allow those assets to be accessed, used and shared by those who could benefit (The Open Data Institute, 2016)

Useful and accessible data infrastructure requires funding to make it sustainable, or the provision of large dynamic datasets to end users is impacted. To help build a picture of weather data infrastructure, we look at each of these areas in turn.

## What data assets are produced in weather data infrastructure?

The core data asset of our global weather data infrastructure is observation data that captures a continuous record of weather and climate data around the world. This includes temperature, rainfall, wind speed and details of a host of other atmospheric, surface and marine conditions.

The collection of observation data is a global effort. The Global Observing System consists of around 11,000 ground-based monitoring stations supplemented with thousands of sensors installed on weather balloons, aircraft and ships.[3] Observations are also collected from a network of radar installations and satellite-based sensors. As we see later, the 'official' observation system is increasingly being supplemented with new sources of observation data from the Internet of Things (IoT).

A combination of historical and current observations is processed using complex numerical models to generate another important set of data assets: detailed weather and climate forecasts. Forecasts are generated at global, regional and national levels with increasing levels of detail to give insight into local conditions.

3 See: 'Observation components of the Global Observing System' at: http://www.wmo.int/pages/prog/www/OSY/Gos-components.html.

The collected observations are analysed and used to help define the starting point for subsequent forecasts. This process is called data assimilation. For early developments in NWP, assimilation was mostly concerned with observations at the main synoptic hours and a snapshot evaluation of the model state was an appropriate approach. Today's observations are dominated by asynoptic satellite and aircraft measurements, where there is a more continuous stream of data. Adding a time dimension to the traditional 3D analysis techniques provides a better way of representing this process.

Forecasts draw on large and complex datasets. There are two types of forecast models: deterministic and probabilistic. Deterministic models produce a single forecast of the most likely weather. Probabilistic models are generated by running a forecast multiple times with different starting conditions to generate multiple projections of possible weather conditions. This is known as the ensemble approach.

Ensemble model forecasts aim to give an indication of the range of possible future states of the atmosphere and oceans (which are a significant driver of the weather in the atmosphere). This overcomes errors introduced by using imperfect measurement of initial starting conditions that are then amplified by the chaotic nature of the atmosphere. Increasing the number of forecast members over a global scale and at higher resolutions result in data volumes increasing exponentially.

The deterministic and ensemble model data is transformed by NMS and private companies into many other data products targeted at specific uses and industries. These include, for example, the weather reports provided to the broadcast media industry, forecasts tailored for the aviation industry or weather warnings for emergency responders.

## Key findings:

- The primary weather data asset is the ongoing observation of our global weather and climate systems from a variety of space, atmospheric, ocean and land-based sources.
- Ensemble models are used to create accurate short, medium and long-term weather and climate forecasts but generate very large data volumes.
- Forecasts are used to create information, products and services that are consumed by the public, media and other sectors.

## What organisations are involved?

Global weather data infrastructure involves a number of public- and private-sector organisations working at different scales of geography.

Created in 1950, The World Meteorological Organisation (WMO) is made up of 191 member states and territories around the world. The WMO was founded on the principle that global coordination was necessary to reap the benefits of weather and climate data. This includes a commitment to weather data and products being freely exchanged around the world (Obasi, 1995).

A similar set of principles underpins the production and use of climate data.[4]

While the WMO has a global outlook, its work is supplemented by regional meteorological organisations like the European Centre for Medium Range Weather Forecasts (ECMWF) and NMSs, such as the Met Office in the UK.

The capabilities and role of NMSs varies considerably around the world. Firstly, while all nations contribute to the gathering of weather observations, individual countries differ in the level of monitoring and sensor infrastructure they have available for collecting data. Secondly, there are differences in the forecasting services provided by different NMSs. While some larger NMSs provide global deterministic and probabilistic forecasts, others focus on collecting observations and offer a more limited range of forecasts. In these cases, leaving others in the value chain to provide additional forecasting data and products.

Gaps in service offerings from NMSs might also be covered by regional meteorological services. For example, projects like the WMO's Global Data-processing and Forecasting system[5] and the ECMWF both provide additional support for countries at the regional and national level.

4 See: 'Global framework for climate services principles' at: http://www.wmo.int/gfcs/principles.
5 See: 'Global data-processing and forecasting system' at: http://public.wmo.int/en/programmes/global-data-processing-and-forecasting-system.

## Key findings:

- There are a variety of global, regional and national meteorological organisations that collectively maintain our global infrastructure and consume its data assets.
- NMSs vary in their responsibilities – some focus purely on their mandate while others offer additional services.
- There are a mix of public- and private-sector organisations involved in the weather data value chain.

## What are the technologies and processes?

Weather data and forecasts need to be delivered to the widest possible audience, in a sustainable way, internationally.

The first level of data sharing – reflecting the first steps in the value chain – involves observation and forecast data being exchanged between meteorological services. National and regional forecasts require the use of data from other geographic areas, making data sharing an important part of the overall system.

Much of the essential data transfer between NMSs currently happens across custom networks, such as the WMO Global Telecommunication System[6] (GTS) which uses dedicated network infrastructure, satellite and radio broadcasts to share observations and forecasts. This networking infrastructure has existed for many years and predates the internet. The WMO is building on its GTS to achieve an overarching WMO Information System[7] (WIS), enabling systematic access, retrieval, dissemination and exchange of data and information of all WMO and related international programmes.

The critical importance of weather data means that the resilience of the overall system is constantly under review. For example, adding or removing individual observation sites may improve or reduce the quality of weather forecasts around the world. Reductions in the number or quality of observations in one country, for example, can impact the quality of forecasts in other countries (Ingleby et al, 2016). Simulations are often carried out to understand the impacts of expected and unexpected changes to the observation network.[8]

6 See: 'Global Telecommunication System' at: http://public.wmo.int/en/programmes/global-telecommunication-system.
7 See: 'WMO Information System' at: http://www.wmo.int/pages/prog/www/WIS.
8 See: 'Observing System Simulation Experiments' at: http://www.met.rdg.ac.uk/~stefano/research/osse/index.html.

In later stages of the weather data value chain, data and information are shared more widely with consumers to comply with legislative requirements, such as the Reuse of Public Sector Information and Competition Law within the UK.

Weather data has traditionally been shared as bulk downloads. Consumers have generally used the File Transfer Protocol (FTP) to download the latest forecasts generated by meteorological services. But these traditional data dissemination methods are becoming unsustainable as datasets produced by NMSs grow exponentially.

There are a number of technologies that can make disseminating large datasets more sustainable.

First, providing weather data via APIs enables NMSs to allow users to be more selective over the data they want, rather than receiving a large data file containing information they may not need. A few NMSs already use APIs as dissemination platforms. For example, the Met Office's DataPoint[9] service provides a variety of APIs that allow application developers to remotely query the Met Office's observation and forecast data. These APIs allow consumers to extract only the data they need for their applications, at the right granularity and in well-defined formats.

Cloud services are increasingly used. For example, the US National Oceanic and Atmospheric Administration (NOAA) have recently started a 'big data project', which uses partnerships with cloud service providers – such as Amazon, Google and Microsoft – to make data available within their platforms.[10] This removes the need for developers to constantly move data between environments.

## Key findings:

- The technical infrastructure for sharing data between NMSs differs from that used to share data more widely.
- Careful resilience planning and monitoring ensures that the observation network remains effective.
- Meteorological services are beginning to embrace new technologies, including APIs and cloud services, in order to better share data with consumers.

9 See: 'Met Office DataPoint' at: http://www.metoffice.gov.uk/datapoint.
10 See: 'Big data project' at: http://www.noaa.gov/big-data-project.

## How is weather data infrastructure evolving?

Based on this summary of the current state of weather data infrastructure, we can reflect on how it is evolving. In particular, it is useful to ask:

- What are the new data assets or sources of data?
- Is there a changing mix of organisations involved in managing global data infrastructure?
- How are the technologies evolving?

These questions are explored in the following sections.

## What are the new sources of weather data?

There are increasing new sources of weather observation data. In recent years, services like Weather Underground and the Met Office's Weather Observation Website have demonstrated the potential for people around the world to contribute weather observations about their local areas – using low-cost home weather stations and sensors, for example. But there is now potential for sensors in cities, homes, cars, cell towers and even mobile phones to contribute observational data that could also be fed into forecast models.

The Weather Signal mobile application[11] is one example of these non-traditional data sources. These new data sources could improve coverage in countries with less-developed infrastructure (Snow, 2013).

This increasing variety of observational data exacerbates the big data challenge. There is increasing volumes of data to be collected, and work is required to understand its quality and role in adding value and insight into the weather data value chain.

11 See: 'App turns smartphone sensors into weather stations' at: https://www.newscientist.com/article/dn23506-app-turns-smartphone-sensors-into-weather-stations

## Key findings:

- There are many new ‘non-traditional’ sources of observation data, e.g. sensors in homes, cities and phones.
- New data sources are producing an increasing volume of data to be collected and processed.
- Data quality is not necessarily increasing, meaning that more analysis is needed to extract value.

## How is the profile of organisations changing?

Given the growth of weather data sources, many new organisations are becoming involved in the weather data value chain. Many are from businesses that are using new technologies or cloud platforms to create new weather data and products.

For example:

- Organisations like PlanetIQ are using new technology to offer commercially licensed weather observations from their own satellite network.
- New types of sensors, such as TAMDAR, offer companies like Panasonic the ability to combine public and private data with custom forecasting models to create new commercial forecasts.[12]
- IBM recently acquired The Weather Company[13] with a goal of combining their weather platform with IBM’s Watson to power a new Internet of Things platform.

Greater access to both open data and tailored, commercial weather data products is enabling a range of new startups and innovators. These are often targeting specific types of users in the value chain. For example, following their acquisition by Monsanto, Climate Corp has been described as the first billion-dollar open data business. As a ‘digital agriculture’ business they are combining weather, soil and field data to provide services to farmers.

12 See: ‘TV maker Panasonic says it has developed the world’s best weather model’ at: http://arstechnica.co.uk/science/2016/04/panasonic-weather-forecasting-model/.

13 See: ‘IBM closes deal to acquire the weather company’s product and technology business’ at: http://www-03.ibm.com/press/us/en/pressrelease/48884.wss.

The increased availability of private sector forecasting and weather products means that some meteorological services are choosing to focus only on their core public task. For example, Met Norway[14] and the Israeli weather service (Meteoworld, 2016) have been embracing open data as a means to enable their data to be widely reused and shared. They are showing that an open approach can deliver cost savings while enabling other organisations in the value chain to provide more services.

Citizens are also changing how they find and use weather data, with an explosion in the adoption of smartphones and demand for timely weather forecasts and warnings.

Google and other search engines have begun to incorporate weather forecasts directly into their search results. However, due to the increasing variety of information providers, it may not be clear to consumers where forecasts and related information are being sourced from.

These issues are creating a number of pressures on existing meteorological services, both in terms of sharing their data with a wider audience, and in maintaining their authoritative voice, as we will discuss in the final section.

### Key findings:

- There are an increasing number of private companies operating at all steps of the value chain.
- Consumers may not always be clear on the source of their weather information.

## How are weather data infrastructure technologies evolving?

A recent article by the WMO highlights that “Global weather prediction is one of the largest scientific and technological challenges in the twenty-first century” (Bauer et al, 2015). Increasing the accuracy of weather forecasting allows individual storms and other weather events to be predicted and tracked more precisely, while improvements in the accuracy of long-range forecasts enables more effective long-term planning.

Improving accuracy, however, requires significant investment and improvement in numerical models and computing infrastructure necessary to run and disseminate their data outputs.

14 See: ‘Met Norway, Data Policy and Data Services’ at: http://met.no/English/Data_Policy_and_Data_Services/.

Running ensemble forecasting models on millions of observations every hour results in larger volumes of data being created. This raises a significant challenge around the collection, production, processing and distribution of weather forecasts.[15]

The Met Office's recently acquired £97m supercomputer provides the hardware infrastructure necessary to run weather forecasts more frequently and at increasing levels of precision.[16] The Met Office is now able to provide seasonal predictions with a 62% accuracy (MacGregor, 2016). Every hourly run of the ensemble model currently generates 400 GB of data, and this is projected to increase to 1TB per hour by 2020.

Growing volumes of forecast data create challenges in how it can be made available to customers. Its sheer volume makes it prohibitive to ship all the raw files to users in a timely manner. Traditional approaches like FTP and bulk downloads do not scale, impacting the sustainability of the weather data infrastructure.

It takes sophisticated analysis and computing infrastructure to draw value from forecast data. Many potential consumers do not have access to the necessary computing resources or expertise required to successfully use the data.

As a result, many meteorological services are exploring cloud infrastructure – like commodity cloud services – to scale out how they process and deliver their forecasting data. But this creates another challenge in that all data that should be made available under mandate need to be transmitted on time and in full.

While running limited area and time weather forecast models (NWP) in the cloud is possible, developing the models themselves is not feasible in a cloud architecture. Also, initialising weather models from observations,[17] long time series or large areal coverage are not practical or affordable to run in existing or projected cloud-based environments. They require highly specialised computer architecture optimised for these tasks. While these activities are expected to become possible using cloud-based architectures in the future, the high performance requirements for timely delivery of NWP suggest that it is not likely to be affordable for a least a decade.

15 See: 'Big Data Drive' at: http://www.metoffice.gov.uk/services/data-provision/big-data-drive.
16 See: '£97m supercomputer makes UK world leader in weather and climate science' at: http://www.metoffice.gov.uk/news/releases/2014/new-hpc.
17 See: 'Met Office Data Assimilation Methods' at: http://www.metoffice.gov.uk/research/weather/data-assimilation/data-assimilation-methods.

NMSs, with global deterministic and ensemble models, have to consider alternatives. For example, they will need to allow users to run their analysis 'on premise', enabling customers to select and pull the data of interest, which may not be available on cloud services. This is known as 'taking the problem to the data'.[18]

The Met Office has recognised the impacts of disseminating large datasets, recently creating a campaign called the 'Big Data Drive'.[19] Its aim is to highlight the big data challenge, and how that challenge can be addressed through the creation of a discoverable and accessible platform using industry open standards (Carne, 2016). This will enable data customers to choose the data they want, in the form they want and at the speed and service level they require.

## Key findings:

- Meteorological services have a big data problem: the increasing volume, variety and velocity of forecasting data mean new dissemination approaches are needed.
- There are substantial costs associated with creating more accurate forecasts and providing a sustainable infrastructure that will allow it to be used by consumers.
- There is a need for more open standards to facilitate data interchange between different organisations in the value chain.

18 See: 'Met Office Big Data Drive' at: http://www.metoffice.gov.uk/services/data-provision/big-data-drive.
19 Ibid

# The future challenges for our weather data infrastructure

In this review we have highlighted several technology trends that are starting to have an impact on our global weather data infrastructure.

- There is an **increasing variety of sources of weather observations** from ground, air, sea and space based monitoring and sensors.
- **Supercomputing infrastructure is enabling new and improved weather models** to create more accurate forecasts, generating rapidly increasing volumes of data. However, the growth of supercomputing in science and business may be limited by the cost considerations of the power required by these systems.
- **Cloud-based platforms are enabling new ways to disseminate and analyse weather data** using APIs and hosted analytical tools.

These trends exist within a wider landscape of innovation and evolving consumer expectations, where instant and real-time access to data is becoming the default.

Collectively these factors are starting to create several impacts on the sector:

- Globally, **commercial organisations are increasingly taking roles across the whole weather data value chain**, creating challenges for national meteorological services in maintaining a visible, authoritative voice.
- **Meteorological services are facing difficult decisions** about how to continue to provide open, sustainable access to weather data, while also continuing to invest in hardware, technology and platforms to meet customer needs.
- **New approaches and platforms are needed to deliver tailored weather products** in a form that can be easily reused by a variety of data consumers.

These impacts raise questions about the future shape of our global weather data infrastructure. We hope that this report will encourage discussion around how that infrastructure should be owned and operated.

# About this report

This report was produced by the Open Data Institute (ODI) in collaboration with the Met Office.

The ODI helps governments and businesses around the world to get data to people who need it. It is independent, nonprofit and nonpartisan, founded in 2012 by Sir Tim Berners-Lee and Sir Nigel Shadbolt. From its headquarters in London and via its global network of startups, members and nodes, the ODI offers training, research and strategic advice for organisations looking to explore the possibilities of data.

The Met Office is the UK's NMS, providing 24x7 world-renowned scientific excellence in weather, climate and environmental forecasts and severe weather warnings for the protection of life and property.

The Met Office has been at the forefront of global weather and climate science for 160 years. Data, both in terms of observations of the weather and the output from forecasting models, are at the heart of its operations. It combines the latest science with ground breaking advances in technology and local understanding to deliver an operational advantage to its customers both at home and overseas. The Met Office's knowledge, experience and flexibility allow it to apply its science across business and government, managing the risks and opportunities from our weather as they arise.

# Bibliography


Atz, U., Fawcett, J. and Hardinges, J., (2015). *How to prioritise open data to drive global development*. London: Open Data Institute

Bauer, P., Thorpe, A., Brunet, G., (2015)  Review: The Quiet Revolution of Numerical Weather Prediction. *Nature*. 64 (2).

Carne, R., (2016) *Open Weather – Weather on the Web., [online]*. Met Office. Available at: http://meetings.wmo.int/CBS-16/_layouts/15/WopiFrame.aspx?sourcedoc=/CBS-16/TECO/Presentations/Background-to-presentations/1C2-TECO_Open_Weather%20(final).docx&action=default [Accessed 11 Apr. 2017].

Callen, J., (2014). *The value of government weather and climate data*. Washington, DC: Economics and Statistics Administration., [online]. Available at: http://www.esa.doc.gov/economic-briefings/value-government-weather-and-climate-data [Accessed 11 Apr. 2017].

Carolan, L., Smith, F., Protonotarios, V., Schaap, B., Broad, E., Hardinges, J., Gerry, W., (2015). *How can we improve agriculture, food and nutrition with open data?* London: Open Data Institute.

Ingleby, B., Rodwell, M., Isaksen, L., (2016) *Impact of a large reduction in the number of Russian radiosonde reports*., [online poster]. Available at: http://www.isda2016.net/abstracts/posters/IsaksenThevalue.html [Accessed 11 Apr. 2017].

MacGregor, A., (2016). *Met Office accurately predicts winter weather a year in advance., [online]*. Available at: https://thestack.com/cloud/2016/10/18/met-office-accurately-predicts-winter-weather-a-year-in-advance [Accessed 10 Apr. 2017].

Meteoworld, (2016). *Case study: Free data access – the experience of the Israel Meteorological Service., [online]*. Available at: http://public.wmo.int/en/resources/meteoworld/case-study-free-data-access-%E2%80%93-experience-of-israel-meteorological-service [Accessed 10 Apr. 2017].

Obasi, G., (1995). *World Meteorological Organization Resolution 40 Cover Letter*., [online]. Available at: http://www.nws.noaa.gov/im/wmocovr.htm [Accessed 10 Apr. 2017].

Open Data Charter, (n.d.). *Agriculture open data package – meteorological data*., [online]. Available at: http://opendatacharter.net/agriculture-open-data-package/section-2-towards-open-data-infrastructure-agriculture/natural-resources-earth-environment-data/meteorological-data [Accessed 10 Apr. 2017].

The Open Data Institute, (2016). *What is data infrastructure?*, [online]. Available at: http://theodi.org/what-is-data-infrastructure [Accessed 10 Apr. 2017].

Snow, J., (2013). *Non-Traditional Approaches to Weather Observations in Developing Countries., [online].* Available at: http://www.ifc.org/wps/wcm/connect/topics_ext_content/ifc_external_corporate_site/cb_home/publications/publication_weatherobservations_landing [Accessed 10 Apr. 2017].

World Meteorological Organization, (2016). *Global Observing System*., [online] Available at: https://public.wmo.int/en/programmes/global-observing-system.